

Europäisches Patentamt
European Patent Office
Office européen des brevets



(11) **EP 1 338 957 A2**

(12) **EUROPEAN PATENT APPLICATION**

(43) Date of publication:
27.08.2003 Bulletin 2003/35

(51) Int Cl.7: **G06F 9/38**

(21) Application number: **03011071.2**

(22) Date of filing: **27.10.1994**

(84) Designated Contracting States:
DE FR GB IT NL

• **Arya, Siamak**
Cupertino, CA 95014 (US)

(30) Priority: **05.11.1993 US 147800**

(74) Representative: **Sparing - Röhl - Henseler**
Patentanwälte
Rethelstrasse 123
40237 Düsseldorf (DE)

(62) Document number(s) of the earlier application(s) in
accordance with Art. 76 EPC:
01101879.3 / 1 102 166
99113656.5 / 0 950 946

Remarks:

This application was filed on 20 - 05 - 2003 as a
divisional application to the application mentioned
under INID code 62.

(71) Applicant: **Intergraph Corporation**
Huntsville, Alabama 35824 (US)

(72) Inventors:
• **Sachs, Howard G.**
Los Altos, CA 94022-2230 (US)

(54) **Software scheduled superscalar computer architecture**

(57) A computing system is described in which groups of individual instructions are executed in parallel by processing pipelines, and instructions to be executed in parallel by different pipelines are supplied to the pipelines simultaneously. During compilation of the instructions those which can be executed in parallel are identified. The system includes a register for storing an arbitrary number of the instructions to be executed. The

instructions to be executed are tagged with pipeline identification tags and group identification tags indicative of the pipeline to which they should be dispatched, and the group of instructions which may be dispatched during the same operation. The pipeline and group identification tags are used to dispatch the appropriate groups of instructions simultaneously to the differing pipelines.

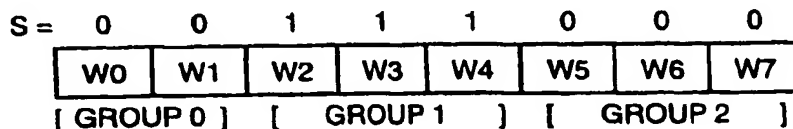


FIG. 5B

EP 1 338 957 A2

Description

[0001] This invention relates to a method according to claim 1; a processor according to claim 15 and an apparatus with a processor according to claim 28, i.e. to the architecture of computing systems, and in particular to an architecture in which groups of instructions may be executed in parallel, as well as to methods and apparatus for accomplishing that.

[0002] A common goal in the design of computer architectures is to increase the speed of execution of a given set of instructions. Many solutions have been proposed for this problem, and these solutions generally can be divided into two groups.

[0003] According to a first approach, the speed of execution of individual instructions is increased by using techniques directed to decreasing the time required to execute a group of instructions serially. Such techniques include employing simple fixed-width instructions, pipelined execution units, separate instruction and data caches, increasing the clock rate of the instruction processor, employing a reduced set of instructions, using branch prediction techniques, and the like. As a result it is now possible to reduce the number of clocks to execute an instruction to approximately one. Thus, in these approaches, the instruction execution rate is limited to the clock speed for the system.

[0004] To push the limits of instruction execution to higher levels, a second approach is to issue more than one instruction per clock cycle, in other words, to issue instructions in parallel. This allows the instruction execution rate to exceed the clock rate. There are two classical approaches to parallel execution of instructions.

[0005] Computing systems that fetch and examine several instructions simultaneously to find parallelism in existing instruction streams to determine if any can be issued together are known as superscaler computing systems. In a conventional superscaler system, a small number of independent instructions are issued in each clock cycle. Techniques are provided, however, to prevent more than one instruction from issuing if the instructions fetched are dependent upon each other or do not meet other special criteria. There is a high hardware overhead associated with this hardware instruction scheduling process. Typical superscaler machines include the Intel i960CA, the IBM RIOS, the Intergraph Clipper C400, the Motorola 88110, the Sun SuperSparc, the Hewlett-Packard PA-RISC 7100, the DEC Alpha, and the Intel Pentium.

[0006] Many researchers have proposed techniques for superscaler multiple instruction issue. Agerwala, T., and J. Cocke [1987] "High Performance Reduced Instruction Set Processors," IBM Tech. Rep. (March), proposed this approach and coined the name "superscaler." IBM described a computing system based on these ideas, and now manufactures and sells that machine as the RS/6000 system. This system is capable of issuing up to four instructions per dock and is described in "The

IBM RISC System/6000 Processor," IBM J. of Res. & Develop. (January, 1990) 34:1.

[0007] The other classical approach to parallel instruction execution is to employ a "wide-word" or "very long instruction word" (VLIW) architecture. A VLIW machine requires a new instruction set architecture with a wide-word format. A VLIW format instruction is a long fixed-width instruction that encodes multiple concurrent operations. VLIW systems use multiple independent functional units. Instead of issuing multiple independent instructions to the units, a VLIW system combines the multiple operations into one very long instruction. For example, in a VLIW system, multiple Integer operations, floating point operations, and memory references may be combined in a single "Instruction." Each VLIW instruction thus includes a set of fields, each of which is interpreted and supplied to an appropriate functional unit. Although the wide-word instructions are fetched and executed sequentially, because each word controls the entire breadth of the parallel execution hardware, highly parallel operation results. Wide-word machines have the advantage of scheduling parallel operation statically, when the instructions are compiled. The fixed width instruction word and its parallel hardware, however, are designed to fit the maximum parallelism that might be available in the code, and most of the time far less parallelism is available in the code. Thus for much of the execution time, most of the instruction bandwidth and the instruction memory are unused.

[0008] There is often a very limited amount of parallelism available in a randomly chosen sequence of instructions, especially if the functional units are pipelined. When the units are pipelined, operations being issued on a given dock cycle cannot depend upon the outcome of any of the previously issued operations already in the pipeline. Thus, to efficiently employ VLIW, many more parallel operations are required than the number of functional units.

[0009] Another disadvantage of VLIW architectures which results from the fixed number of slots in the very long instruction word for classes of instructions, is that a typical VLIW instruction will contain information in only a few of its fields. This is inefficient requiring the system to be designed for a circumstance that occurs only rarely - a fully populated instruction word.

[0010] Another disadvantage of VLIW systems is the need to increase the amount of code. Whenever an instruction is not full, the unused functional units translate to wasted bits, no-ops, in the instruction coding. Thus useful memory and/or instruction cache space is filled with useless no-op instructions. In short, VLIW machines tend to be wasteful of memory space and memory bandwidth except for only a very limited class of programs.

[0011] The term VLIW was coined by J. A. Fisher and his colleagues in Fisher, J. A., J. R. Ellis, J. C. Ruttenberg, and A. Nicolau [1984], "Parallel Processing: A Smart Compiler and a Dumb Machine," Proc. SIGPLAN

Conf. on Compiler Construction (June), Palo Alto, CA, 11-16. Such a machine was commercialized by Multiflow Corporation.

[0012] For a more detailed description of both superscaler and VLIW architectures, see Computer Architecture - a Quantitative Approach, John L. Hennessy and David A. Patterson, Morgan Kaufmann Publishers, 1990.

[0013] The invention is defined in claims 1, 15 and 28.

[0014] We have developed a computing system architecture, which we term software-scheduled superscaler, which enables instructions to be executed both sequentially and in parallel, yet without wasting space in the instruction cache or registers. Like a wide-word machine, we provide for static scheduling of concurrent operations at program compilation. Instructions are also stored and loaded into fixed width frames (equal to the width of a cache line). Like a superscaler machine, however, we employ a traditional instruction set, in which each instruction encodes only one basic operation (load, store, etc.). We achieve concurrence by fetching and dispatching "groups" of simple individual instructions, arranged in any order. The architecture of our invention relies upon the compiler to assign instruction sequence codes to individual instructions at the time they are compiled. During execution these instruction sequence codes are used to sort the instructions into appropriate groups and execute them in the desired order. Thus our architecture does not suffer the high hardware overhead and runtime constraints of the superscaler strategy, nor does it suffer the wasted instruction bandwidth and memory typical of VLIW systems.

[0015] Our system includes a mechanism, an associative crossbar, which routes in parallel each instruction in an arbitrarily selected group to an appropriate pipeline, as determined by a pipeline tag applied to that instruction during compilation. Preferably, the pipeline tag will correspond to the type of functional unit required for execution of that instruction, e.g., floating point unit 1. All instructions in a selected group can be dispatched simultaneously.

[0016] Thus, in one implementation, our system includes a cache line, register, or other means for holding at least one group of instructions to be executed in parallel, each instruction in the group having associated therewith a pipeline identifier indicative of the pipeline for executing that instruction and a group identifier indicative of the group of instructions to be executed in parallel. The group identifier causes all instructions having the same group identifier to be executed simultaneously, while the pipeline identifier causes individual instructions in the group to be supplied to an appropriate pipeline.

[0017] In another embodiment the register holds multiple groups of instructions, and all of the instructions in each group having a common group identifier are placed next to each other, with the group of instructions to be executed first placed at one end of the register, and the

instructions in the group to be executed last placed at the other end of the register.

[0018] In another embodiment of our invention a method of executing arbitrary numbers of instructions in a stream of instructions in parallel includes the steps of compiling the instructions to determine which instructions can be executed simultaneously, assigning group identifiers to sets of instructions that can be executed in parallel, determining a pipeline for execution of each instruction, assigning a pipeline identifier to each instruction, and placing the instructions in a cache line or register for execution by the pipelines.

[0019] Further embodiments of the invention are:

1. A computing system for executing groups of individual instructions in parallel by processing pipelines, the system comprising:

storage means for holding at least one group of instructions to be executed in parallel, each instruction in the group having associated therewith a pipeline identifier indicative of the pipeline for executing that instruction and a group identifier indicative of the group of instructions to be executed in parallel;
means responsive to the group identifier for causing all instructions having the same group identifier to be executed at the same time; and
means responsive to the pipeline identifier of the individual instructions in the group to supply each instruction in the group to be executed in parallel to an appropriate pipeline.

2. A computing system as in embodiment 1 wherein the storage means includes the at least one group of instructions, and for each instruction the storage means includes the group identifier and the pipeline identifier.

3. A computing system as in embodiment 2 wherein each instruction in the at least one group of instructions has associated therewith a different pipeline identifier.

4. A computing system as in embodiment 1 wherein the storage means holds at least two groups of instructions, and all of the instructions in each group having associated therewith a common group identifier are placed adjacent to each other in the storage means.

5. A computing system as in embodiment 4 wherein:

the storage means comprises a line in a cache memory having a fixed number of storage locations;
the group of instructions to be executed first is placed at one end of the line in the cache memory, and the instructions in the group to be executed last is placed at the other end of the line in the cache memory.

6. A method of executing arbitrary numbers of instructions in a stream of instructions in parallel which have been compiled to determine which instructions can be executed at the same time, the method comprising:

in response to the compilation assigning group identifiers to sets of instructions which can be executed in parallel;
determining a pipeline for execution of each instruction in a group to be executed;
assigning a pipeline identifier to each instruction in the group; and
placing the instructions in a register for execution by the pipelines.

7. A method as in embodiment 6 further comprising the step of executing a group of instructions in parallel.

8. A method as in embodiment 7 wherein the register holds at least two groups of instructions, and the step of placing the instructions in the register for execution by the pipelines comprises placing the instructions in each group having associated therewith a common group identifier adjacent to each other in the register.

9. A method as in embodiment 8 the step of executing a group of instructions in parallel comprises coupling the register to detection means to receive the group identifier of each instruction in the register and the group identifier of the next group of instructions to be supplied to the pipelines; and

supplying only the instructions with the next group identifier to the pipeline execution units.

10. In a computing system in which groups of individual instructions are executable in parallel by processing pipelines, a method for supplying each instruction in a group to be executed in parallel to an appropriate pipeline, the method comprising:

storing in storage an instruction frame, the frame including at least one group of instructions to be executed in parallel, each instruction in the group having associated therewith a pipeline identifier indicative of the pipeline which will execute that instruction and a group identifier indicative of the group identification; comparing the group identifier of each instruction in the instruction frame and a group identifier of those instructions to be next executed in parallel; and

using the pipeline identifier of those instructions to be next executed in parallel to control an execution unit to execute all of the instructions in the group in separate pipelines.

11. In a computing system in which groups of individual instructions are executable in parallel by

processing pipelines, apparatus for routing each instruction in a group to be executed in parallel to an appropriate pipeline, the apparatus comprising:

storage for holding at least one group of instructions to be executed in parallel, each instruction in the group having associated therewith a pipeline identifier indicative of the pipeline for executing that instruction and a group identifier to designate among the instructions present in the storage those instructions which may be simultaneously supplied to the processing pipelines;

a crossbar having a first set of connectors coupled to the storage for receiving instructions therefrom and a second set of connectors coupled to the processing pipelines;

means responsive to the pipeline identifier of the individual instructions in the group for routing individual instructions onto appropriate ones of the second set of connectors, to thereby supply each instruction in the group to be executed in parallel to the appropriate pipeline.

12. Apparatus as in embodiment 11 wherein:

the first set of connectors consists of a set of first communication buses, one for each instruction in the storage;

the second set of connectors consists of a set of second communication buses, one for each pipeline; and the means responsive to the pipeline identifier comprises:

a set of decoders coupled to the storage to receive as first input signals the pipeline identifiers and in response thereto supply as output signals a switch control signal; and

a set of switches, coupled to the decoders, one switch at the intersection of each of the first set of connectors with the second set of connectors, the switches providing connections in response to receiving the switch control signal to thereby supply each instruction in the group to be executed in parallel to the appropriate pipeline.

13. Apparatus as in embodiment 12 further comprising:

detection means coupled to receive the group identifier of each instruction in the storage and connected to receive information regarding the group identifier of the next group of instructions to be supplied to the pipelines, and in response thereto supply a group control signal; and

wherein the set of decoders coupled to the storage are also coupled to the detection means to receive the group control signal and in response thereto supplies a switch control signal for only those instructions in the group to be supplied to the pipelines.

14. Apparatus as in embodiment 13 wherein the detection means comprises a multiplexer coupled to receive each of the group identifiers of instructions in the storage and compare them to the information regarding the group identifier of the next group of instructions to be supplied to the pipelines.

15. Apparatus as in embodiment 14 wherein the multiplexer supplies an output signal to the decoders to indicate the group identifier of the next group of instructions to be supplied to the pipelines.

16. In a computing system in which groups of individual instructions are executable in parallel by processing pipelines, apparatus for routing each instruction in a group to be executed in parallel to an appropriate pipeline, the apparatus comprising:

a storage for holding an instruction frame, the frame including at least one group of instructions to be executed in parallel, each instruction in the group having associated therewith a pipeline identifier indicative of the pipeline to which that instruction is to be issued and a group identifier indicative of the group identification;

a crossbar switch having a first set of connectors coupled to the storage for receiving instructions therefrom and a second set of connectors coupled to the processing pipelines;

selection means connected to receive the group identification of each instruction in the instruction frame and connected to receive information about the group identifier of those instructions to be next executed in parallel for supplying in response thereto an output signal indicative of the next set of instructions to be executed in parallel; and

decoder means coupled to receive the output signal and each of the pipeline identifiers of the instructions in the storage for selectively connecting ones of the first set of connectors to ones of the second set of connectors to thereby supply each instruction in the group to be executed in parallel to the appropriate pipeline.

17. Apparatus as in embodiment 16 wherein the first set of connectors consists of a set of first communication buses, one for each instruction in the storage;

the second set of connectors consists of a set of second communication buses, one for each pipeline;

the decoder means comprises a set of decod-

ers coupled to receive as first input signals the pipeline identifiers and as second input signals information about the group identifier of the next group of instructions to be executed by the pipelines and in response thereto supply as output signals a switch control signal; and

the crossbar switch includes a set of switches, one at the intersection of each of the first set of connectors with the second set of connectors, the switches providing connections in response to receiving the switch control signal to thereby supply each instruction in the group to be executed in parallel to the appropriate pipeline.

18. Apparatus as in embodiment 17 wherein the selection means coupled to the storage comprises a multiplexer coupled to receive each of the group identifiers of instructions in the storage and compare them to information regarding the group identifier of the next group of instructions to be supplied to the pipelines.

19. Apparatus as in embodiment 18 wherein the multiplexer supplies an output signal to the decoders to select the group identifier of the next group of instructions to be supplied to the pipelines.

20. In a computing system in which groups of individual instructions are executable in parallel by processing pipelines, a method for transferring each instruction in a group to be executed through a crossbar switch having a first set of connectors coupled to the storage for receiving instructions therefrom and a second set of connectors coupled to the processing pipelines, the method comprising: storing in storage at least one group of instructions to be executed in parallel, each instruction in the group having associated therewith a pipeline identifier indicative of the pipeline which will execute that instruction; and

using the pipeline identifiers of the individual instructions in the at least one group of instructions which are to be executed next to control switches between the first set of connectors and the second set of connectors to thereby supply each instruction in the group to be executed in parallel to the appropriate pipeline.

21. A method as in embodiment 20 wherein the step of using comprises:

supplying the pipeline identifiers of the individual instructions in the at least one group of instructions to a corresponding number of decoders, each of which provides an output signal indicative of the pipeline identifiers; and using the decoder output signals to control the switches between the first set of connectors and the second set of connectors to thereby supply each instruction in the group to be executed in parallel to the appropriate pipeline.

22. A method as in embodiment 21 wherein each of the instructions in the storage further includes a group identifier to designate among the instructions present in the storage which may be simultaneously supplied to the processing pipelines, and the method further comprises:

supplying information about the group identifier of the next group of instructions to be executed by the pipelines together with the group identifiers of the individual instructions in the at least one group of instructions to a selector; comparing the group identifier of the next group of instructions to be executed by the pipelines with the group identifiers of the individual instructions in the at least one group of instructions, to provide output comparison signals; and using both the output comparison signals and the decoder output signals to control the switches between the first set of connectors and the second set of connectors to thereby supply each instruction in the group to be executed in parallel to the appropriate pipeline.

23. In a computing system in which groups of individual instructions are executable in parallel by processing pipelines, a method for supplying each instruction in a group to be executed in parallel to an appropriate pipeline, the method comprising:

storing in storage an instruction frame, the frame including at least one group of instructions to be executed in parallel, each instruction in the group having associated therewith a pipeline identifier indicative of the pipeline which will execute that instruction and a group identifier indicative of the group identification; comparing the group identifier of each instruction in the instruction frame and a group identifier of those instructions to be next executed in parallel; and using the pipeline identifier of those instructions to be next executed in parallel to control switches in a crossbar switch having a first set of connectors coupled to the storage for receiving instructions therefrom and a second set of connectors coupled to the processing pipelines to thereby supply each instruction in the group to be executed in parallel to the appropriate pipeline.

24. A computing system for executing groups of individual instructions in parallel by processing pipelines, the system comprising:

storage means for holding at least one group of instructions to be executed in parallel, each in-

struction in the group having associated therewith a pipeline identifier indicative of the pipeline for executing that instruction and a group identifier indicative of a group of instructions to be executed in parallel;

means responsive to the group identifier for causing all instructions having the same group identifier to be executed at the same time; and means responsive to the pipeline identifier of the individual instructions in the group to supply each instruction in the group to be executed in parallel to an appropriate pipeline,

wherein the storage means includes the at least one group of instructions, and for each instruction the storage means includes the group identifier and the pipeline identifier.

25. The system of embodiment 24, wherein each instruction in the at least one group of instructions has associated therewith a different pipeline identifier.

26. The system of embodiment 24 or 25, wherein the storage means holds at least two groups of instructions, and all of the instructions in each group having associated therewith a common group identifier are placed adjacent to each other in the storage means.

27. The system of embodiment 26, wherein the storage means comprises a line in a cache memory having a fixed number of storage locations; the group of instructions to be executed first is placed at one end of the line in the cache memory, and the instructions in the group to be executed last is placed at the other end of the line in the cache memory.

28. The system of any one of the embodiments 24 to 27, wherein a crossbar having a first set of connectors coupled to the storage means for receiving instructions therefrom and a second set of connectors coupled to the processing pipelines is provided.

29. The system of embodiment 28, wherein the first and the second set of connectors, respectively, consist of a set of communication buses, one for each instruction in the storage means and one for each pipeline, respectively, and wherein a set of decoders coupled to the storage means to receive as first input signals the pipeline identifiers and in response thereto supply as output signals a switch control signal and a set of switches, coupled to the decoders with one switch at the intersection of each of the first set of connectors, are provided, the switches providing connections in response to receiving the switch control signal to thereby supply each instruction in the group to be executed in parallel to the appropriate pipeline.

30. The system of embodiment 29, further comprising detection means coupled to receive the group identifier of each construction in the storage means

and connected to receive information regarding the group identifier of the next group of instructions to be supplied to the pipelines, and in response thereto supply a group control signal, and wherein the set of decoders coupled to the storage means are also coupled to the detection means to receive the group control signal and in response thereto supplies a switch control signal for only those instructions in the group to be supplied to the pipelines.

31. The system of embodiment 30, wherein the detection means comprises a multiplexer coupled to receive each of the group of identifiers of instructions in the storage means and compare them to the information regarding the group identifier of the next group of instructions to be supplied to the pipelines.

32. The system of embodiment 31, wherein the multiplexer supplies an output signal to the decoders to indicate the group identifier of the next group of instructions to be supplied to the pipelines.

33. The system of any one of the embodiments 24 to 32, wherein the group of instructions is an instruction frame.

34. The system of embodiment 33, further comprising selection means connected to receive the group identifier of each instruction in the instruction frame and connected to receive information about the group identifier of those instructions to be next executed in parallel for supplying in response thereto an output signal indicative of the next set of instructions to be executed in parallel, and decoder means coupled to receive the output signal and each of the pipeline identifiers of the instructions in the storage for selectively connecting ones of the first set of connectors to thereby supply each instruction in the group to be executed in parallel at the appropriate pipeline.

35. A method for executing groups of individual instructions in parallel by processing pipelines, the method comprising:

storing in storage means for holding at least one group of instructions to be executed in parallel, each instruction in the group having associated therewith a pipeline identifier indicative of the pipeline for executing that instruction and a group identifier indicative of a group of instructions to be executed in parallel;
comparing the group identifier of each instruction in the group of instructions and a group identifier of those instructions to be next executed in parallel; and
using the pipeline identifier of those instructions to be next executed in parallel to control an execution unit to execute all of the instructions in the group in separate pipelines.

36. The method of embodiment 35, further compris-

ing transferring each instruction in a group to be executed through a crossbar having a first set of connectors coupled to the storage means for receiving instructions therefrom and a second set of connectors coupled to the processing pipelines and using the pipeline identifiers of the individual instructions in the at least one group of instructions which are to be executed next to control switches between the first set of connectors and the second set of connectors to thereby supply each instruction in the group to be executed in parallel to the appropriate pipeline.

37. The method of embodiment 36, wherein the step of using comprises supplying the pipeline identifiers of the individual instructions in the at least one group of instructions to a corresponding number of decoders, each of which provides an output signal indicative of the pipeline identifiers; and using the decoder output signals to control the switches between the first set of connectors and the second set of connectors.

38. A computing system in which groups of instructions are issued in parallel to processing pipelines (0, ..., 7), the computing system comprising:

storage means (50; 70, 74) for holding an instruction frame, the instruction frame including a plurality of instructions including at least one group of instructions, the instructions in the at least one group of instructions to be issued in parallel, the at least one group being determined at compile time,

wherein the instruction frame including data associated with the plurality of instructions in the instruction frame, the data indicative of which instructions are included in the one group of instructions and further indicative of processing pipelines (0, ..., 7) appropriate for the plurality of instructions in the instruction frame; and by

means (60; 100) responsive to the data associated with the plurality of instructions for issuing the instructions in the one group of instructions in parallel to processing pipelines (0, ..., 7) appropriate for the instructions in the one group of instructions.

39. The computing system of embodiment 38, wherein at least one group of instructions comprises at least two or at least three instructions.

40. The computing system of embodiment 38 or 39, wherein the plurality of instructions also includes at least another instruction belonging to another group of instructions, the another instruction to be issued after the at least one group of instructions has been issued.

41. The computing system of one of the embodiments 38 to 40, wherein the processing pipelines (0, ..., 7) appropriate for the instructions in the one group of instructions are respectively coupled to ex-

ecution units appropriate for the instructions in the one group of instructions.

42. The computing system of one of the embodiments 38 to 41, wherein an execution unit appropriate for one instruction in the one group of instructions is a memory or an arithmetic logic unit or a floating point unit or a branch unit.

43. The computing system of one of the embodiments 38 to 42, wherein an execution unit appropriate for a first instruction in the one group of instructions and an execution unit appropriate for a second instruction in the one group of instructions are similar.

44. The computing system of one of the embodiments 38 to 43, wherein the plurality of instructions in the instruction frame are determined at compile time.

45. A method for issuing groups of instructions in parallel to processing pipelines (0, ..., 7), the method comprising:

storing in a storage means (50; 70, 74) an instruction frame, the instruction frame including a plurality of instructions including at least one group of instructions, the instructions in the at least one group of instructions to be issued in parallel, the at least one group being determined at compile time,

wherein the instruction frame includes data associated with the plurality of instructions in the instruction frame, the data indicative of which instructions are included in the one group of instructions and further indicative of processing pipelines (0, ..., 7) appropriate for the plurality of instructions in the instruction frame; and

the one group of instructions is issued in parallel to processing pipelines (0, ..., 7) appropriate for the instructions in the one group of instructions in response to the data associated with the plurality of instructions in the instruction frame.

46. The method of embodiment 45, further comprising a compilation step, during which the plurality of instructions in the instructions frame and the data associated with the plurality of instructions in the instruction frame is determined.

47. The method of embodiment 45 or 46, wherein another instruction of the frame is issued to a processing pipeline (0, ..., 7) appropriate for the another instruction in response to the data associated with the plurality of instructions in the instruction frame following the step of issuing the one group of instructions.

48. The method of one of the embodiments 45 to 47, wherein an instruction in the group is issued to an arithmetic logic unit or a floating point unit or a memory unit.

49. The method of embodiment 48, wherein another

instruction in the group is issued to an arithmetic logic unit or a floating point unit or a memory unit.

50. A processor for operating a computing system according to the method of one of the embodiments 45 to 49, wherein

at least two processing pipelines (0, ..., 7);

storage means (50; 70, 74) for holding at least one group of instructions to be issued in parallel, and for holding data associated with the one group of instructions, the data indicative of processing pipelines (0, ..., 7) appropriate for the instructions in the at least one group of instructions; and

means (60; 100) responsive to the data for issuing the instructions in the one group of instructions in parallel to processing pipelines (0, ..., 7) appropriate for the instructions in the one group of instructions, are provided for.

51. A processor comprising:

a register file having a plurality of registers; an instruction set including instructions (INSTRUCTION) which address the registers, each instruction (INSTRUCTION) being one of a plurality of instruction types;

a plurality of execution units (0...7), each execution unit (0...7) being one of a plurality of types, wherein each instruction type is executed on one or more execution unit types;

and further wherein the instructions (INSTRUCTION) are encoded in frames, each frame including a plurality of instructions (INSTRUCTION) and template bits (S,P) grouped together in an N-bit field, the instructions (INSTRUCTION) being located in instructions slots of the N-bit field, the template bits (S,P) specifying a mapping of the instruction slots to the execution unit types.

52. The processor of embodiment 51 wherein the template bits further specify instruction group boundaries within the frame, with an instruction group comprising a set of statically contiguous instructions that are executed concurrently.

53. The processor of embodiment 51 or 52 wherein the instruction types include integer arithmetic logic unit, memory, floating-point, and branch instructions.

54. The processor of one of the embodiments 51 to 53 wherein the execution unit types include integer, memory, floating-point, and branch execution units.

55. The processor of one of the embodiments 51 to 54 wherein the frame further includes a stop-bit that specifies an inter-frame instruction group boundary.

56. The processor according to one of the embodiments 51 to 55, further comprising a memory that stores the frames, a byte order of the frames in the memory being in a little-endian format or in a big-endian format.

57. The processor according to one of the embodiments 51 to 56 wherein the frames are ordered in the memory from a lowest to a highest memory address.

58. The processor according to one of the embodiments 51 to 57 wherein an instruction in the frame with the lowest memory address precedes an instruction in the frame with the highest memory address.

59. The processor according to one of the embodiments 51 to 58 wherein the template bits are partly determined by hardware.

60. The processor according to one of the embodiments 51 to 59 wherein the template bits are at least partly determined at a compile time.

61. A method for operating a processor comprising:

storing a frame of instructions (INSTRUCTION) in a memory, the frame including a plurality of instructions and template bits (S, P), the plurality of instructions (INSTRUCTIONS) and the template bits (S, P) grouped together in an N-bit field, the instructions (INSTRUCTION) being located in instruction slots of the N-bit field, the template bits (S,P) specifying a mapping of the instruction slots to execution unit types, each Instruction (INSTRUCTION) being one of a plurality of instruction types from an instruction set which address a plurality of registers in a register file; and
executing each instruction type on an execution unit (0...7) from a plurality of execution units (0...7) being one of a plurality of execution unit types,

62. The method according to embodiment 61 comprising the step of specifying instruction group boundaries within the frame at the template bits, with an instruction group comprising a set of statically contiguous instructions that are executed concurrently.

63. The method according to embodiment 61 or 62 comprising the step of including a stop-bit into the frame, said stop-bit specifying an inter-frame instruction group boundary.

64. The method according to embodiment 63 comprising the step of specifying the instruction group boundary to occur after a last instruction of a current frame if the stop-bit is in a first condition.

65. The method according to embodiment 64 comprising the step of specifying the instruction group to include the last instruction the current frame if the stop-bit is in a second condition.

66. The method according to one of the embodiments 61 to 65 comprising the step of storing the frames in a memory with a byte order in a little-endian or a big-endian format.

67. The method according to one of the embodi-

ments 61 to 66 comprising the step of ordering the frames in the memory from a lowest to a highest memory address.

68. The method according to one of the claims 61 to 67 wherein the instruction types include integer arithmetic logic unit, memory, floating point, and branch units.

69. The method according to one of the embodiments 61 to 68 comprising the step of determining the template bits partly by hardware.

70. The method according to one of the embodiments 61 to 69 comprising the step of determining the template bits at least partly at compile time.

71. A memory comprising:

a frame of instructions (INSTRUCTION), the frame including a plurality of instructions (INSTRUCTION) and template bits (S,P), the plurality of instructions (INSTRUCTION) and the template bits (S,P) grouped together in an N-bit field, the instructions (INSTRUCTION) being located in instruction slots of the N-bit field, the template bits (S, P) specifying a mapping of the instruction slots to execution unit types, each instruction (INSTRUCTION) being one of a plurality of instruction types from an instruction set and which address a plurality of registers in a register file,

wherein each instruction type is to be executed on an execution unit (0...7) from a plurality of execution units (0...7) being one of a plurality of execution unit types, and

wherein instructions in the frame may be issued to respective execution units (0...7) at different times.

72. The memory of embodiment 71 wherein the template bits further specify instruction group boundaries within the frame, with an instruction group comprising a set of statically contiguous instructions that are executed concurrently.

73. The memory of embodiment 71 or 72 wherein the frame further includes a stop-bit that specifies an inter-frame instruction group boundary.

74. The memory of embodiment 73 wherein, if the stop bit is in a first condition, the instruction group includes the last instruction of the current frame.

75. The memory according to one of the embodiments 71 to 74 wherein a byte order of the frames stored therein is in a little-endian format or in a big-endian format.

76. The memory according to one of the embodiments 71 to 75 wherein the frames are ordered from a lowest to a highest memory address.

77. The memory according to one of the embodiments 71 to 76 wherein an Instruction (INSTRUCTION) in the frame with the lowest memory address precedes an instruction in the frame with the highest

memory address.

78. The memory according to one of the embodiments 71 to 77, further comprising an unused encoding of the template bits being available for use in a future extension.

79. The memory according to one of the embodiments 71 to 78 wherein the template bits are derivable from machine code.

80. The memory according to one of the embodiments 71 to 79 wherein the template bits are determinable at least partly at a compile time.

Figure 1 is a block diagram illustrating a preferred implementation of this invention;

Figure 2 is a diagram illustrating the data structure of an instruction word in this system;

Figure 3 is a diagram illustrating a group of instruction words;

Figure 4 is a diagram illustrating a frame containing from one to eight groups of instructions;

Figure 5a illustrates the frame structure for one maximum-sized group of eight instructions;

Figure 5b illustrates the frame structure for a typical mix of three intermediate sized groups of instructions;

Figure 5c illustrates the frame structure for eight minimum-sized groups, each of one instruction;

Figure 6 illustrates an instruction word after predecoding;

Figure 7 illustrates the operation of the predecoder;

Figure 8 is a diagram illustrating the overall structure of the instruction cache;

Figure 9 is diagram illustrating the manner in which frames are selected from the instruction cache;

Figure 10 is a diagram illustrating the group selection function in the associative crossbar;

Figure 11 is a diagram illustrating the group dispatch function in the associative crossbar;

Figure 12 is a diagram illustrating a hypothetical frame of instructions; and

Figure 13 is a diagram illustrating the manner in which the groups of instructions in Figure 12 are issued on different clock cycles.

Figure 14 is a diagram illustrating another embodiment of the associative crossbar.

Figure 15 is a diagram illustrating the group select function in further detail.

[0020] Figure 1 is a block diagram of a computer system according to the preferred embodiment of this invention. Figure 1 illustrates the organization of the integrated circuit chips by which the computing system is formed. As depicted, the system includes a first integrated circuit 10 that includes a central processing unit, a floating point unit, and an instruction cache.

[0021] In the preferred embodiment the instruction cache is a 16 kilobyte two-way set-associative 32 byte line cache. A set associative cache is one in which the

lines (or blocks) can be placed only in a restricted set of locations. The line is first mapped into a set, but can be placed anywhere within that set. In a two-way set associative cache, two sets, or compartments, are provided, and each line can be placed in one compartment or the other.

[0022] The system also includes a data cache chip 20 that comprises a 32 kilobyte four-way set-associative 32 byte line cache. The third chip 30 of the system includes a predecoder, a cache controller, and a memory controller. The predecoder and instruction cache are explained further below. For the purposes of this invention, the CPU, FPU, data cache, cache controller and memory controller all may be considered of conventional design.

[0023] The communication paths among the chips are illustrated by arrows in Figure 1. As shown, the CPU/FPU and instruction cache chip communicates over a 32 bit wide bus 12 with the predecoder chip 30. The asterisk is used to indicate that these communications are multiplexed so that a 64 bit word is communicated in two cycles. Chip 10 also receives information over 64 bit wide buses 14, 16 from the data cache 20, and supplies information to the data cache 20 over three 32 bit wide buses 18.

[0024] The specific functions of the predecoder are described in much greater detail below; however, essentially it functions to decode a 32 bit instruction received from the secondary cache into a 64 bit word, and to supply that 64 bit word to the instruction cache on chip 10.

[0025] The cache controller on chip 30 is activated whenever a first level cache miss occurs. Then the cache controller either goes to main memory or to the secondary cache to fetch the needed information. In the preferred embodiment the secondary cache lines are 32 bytes and the cache has an 8 kilobyte page size.

[0026] The data cache chip 20 communicates with the cache controller chip 30 over another 32 bit wide bus. In addition, the cache controller chip 30 communicates over a 64 bit wide bus 32 with the DRAM memory, over a 128 bit wide bus 34 with a secondary cache, and over a 64 bit wide bus 36 to input/output devices.

[0027] As will be described, the system shown in Figure 1 includes both conventional and novel features. The system includes multiple pipelines able to operate in parallel on separate instructions. The instructions that can be dispatched to these parallel pipelines simultaneously, in what we term "instruction groups", have been identified by the compiler and tagged with a group identification tag. Thus, the group tag designates instructions that can be executed simultaneously. Instructions within the group are also tagged with a pipeline tag indicative of the specific pipeline to which that instruction should be dispatched. This operation is also performed by the compiler.

[0028] In this system, each group of instructions can contain an arbitrary number of instructions ordered in an arbitrary sequence. The only limitation is that all in-

structions in the group must be capable of simultaneous execution; e.g. there cannot be data dependency between instructions. The instruction groups are collected into larger sets and are organized into fixed width "frames" and stored. Each frame can contain a variable number of tightly packed instruction groups, depending upon the number of instructions in each group and on the width of the frame.

[0029] Below we describe this concept more fully, as well as describe a mechanism to route in parallel each instruction in an arbitrary selected group to its appropriate pipeline, as determined by the pipeline tag of the instruction.

[0030] In the following description of the word, group, and frame concepts mentioned above, specific bit and byte width are used for the word, group and frame. It should be appreciated that these width are arbitrary, and can be varied as desired. None of the general mechanisms described for achieving the result of this invention depends upon the specific implementation.

[0031] In one embodiment of this system the central processing unit includes eight functional units and is capable of executing eight instructions in parallel. We designate these pipelines using the digits 0 to 7. Also, for this explanation each instruction word is 32 bits (4 bytes) long, with a bit, for example, the high order bit S being reserved as a flag for group identification. Figure 2 therefore shows the general format of all instructions. As shown by Figure 2, bits 0 to 30 represent the instruction, with the high order bit 31 reserved to flag groups of instructions, i.e., collections of instructions the compiler has determined may be executed in parallel.

[0032] Figure 3 illustrates a group of instructions. A group of instructions consists of one to eight instructions (because there are eight pipelines in the preferred implementation) ordered in any arbitrary sequence; each of which can be dispatched to a different parallel pipeline simultaneously.

[0033] Figure 4 illustrates the structure of an instruction frame. In the preferred embodiment an instruction frame is 32 bytes wide and can contain up to eight instruction groups, each comprising from one to eight instructions. This is explained further below.

[0034] When the instruction stream is compiled before execution, the compiler places instructions in the same group next to each other in any order within the group and then places that group in the frame. The instruction groups are ordered within the frame from left to right according to their issue sequence. That is, of the groups of instruction in the frame, the first group to issue is placed in the leftmost position, the second group to issue is placed in the next position to the right, etc. Thus, the last group of instructions to issue within that frame will be placed in the rightmost location in the frame. As explained, the group affiliation of all instructions in the same group is indicated by setting the S bit (bit 31 in Figure 2) to the same value. This value toggles back and forth from 0 to 1 to 0, etc. between adjacent groups

to thereby identify the groups. Thus, all instructions in the first group in a frame have the S bit set to 0, all instructions in the second group have the S bit set to 1, all instructions in the third group have the S bit set to 0, etc. for all groups of instructions in the frame.

[0035] To clarify the use of a frame, Figure 5 illustrates three different frame structures for different hypothetical groups of instructions. In Figure 5a the frame structure for a group of eight instructions, all of which can be issued simultaneously, is shown. The instruction words are designated WO, W1, ..., W7. The S bit for each one of the instruction words has been set to 0 by the compiler, thereby indicating that all eight instructions can be issued simultaneously.

[0036] Figure 5 illustrates the frame structure for a typical mixture of three intermediate sized groups of instructions. In Figure 5b three groups of instructions are designated Group 0, Group 1 and Group 2. Shown at the left-hand side of Figure 5b is Group 0 that consists of two instruction words WO and W1. The S bit for each of these instructions has been set to 0. Group 1 of instructions consists of three instruction words, W2, W3 and W4, each having the S bit set to 1. Finally, Group 2 consists of three instruction words, W5, W6 and W7, each having its S bit set to 0.

[0037] Figure 5c illustrates the frame structure for eight minimum sized groups, each consisting of a single instruction. Because each "group" of a single instruction must be issued before the next group, the S bits toggle in a sequence 01010101 as shown.

[0038] As briefly mentioned above, in the preferred embodiment the group identifiers are associated with individual instructions in a group during compilation. In the preferred embodiment, this is achieved by compiling the instructions to be executed using a well-known compiler technology. During the compilation, the instructions are checked for data dependencies, dependence upon previous branch instructions, or other conditions that preclude their execution in parallel with other instructions. These steps are performed using a well-known compiler. The result of the compilation is a group identifier being associated with each instruction. It is not necessary that the group identifier be added to the instruction as a tag, as shown in the preferred embodiment and described further below. In an alternative approach, the group identifier is provided as a separate tag that is later associated with the instruction. This makes possible the execution of programs on our system, without need to revise the word width.

[0039] In addition, in some embodiments the compiler will determine the appropriate pipeline for execution of an individual instruction. This determination is essentially a determination of the type of instruction provided. For example, load instructions will be sent to the load pipeline, store instructions to the store pipeline, etc. The association of the instruction with the given pipeline can be achieved either by the compiler, or by later examination of the instruction itself, for example during predecoding.

[0040] Referring again to Figure 1, in normal operation the CPU will execute instructions from the instruction cache, according to well-known principles. On an instruction cache miss, however, the entire frame containing the instruction missed is transferred from the main memory into the secondary cache and then into the primary instruction cache, or from the secondary cache to the primary instruction cache, where it occupies one line of the instruction cache memory. Because instructions are only executed out of the instruction cache, all instructions ultimately undergo the following procedure.

[0041] At the time a frame is transferred into the instruction cache, the instruction word in that frame is predecoded by the predecoder 30 (Figure 1), which as is explained below decodes the retrieved instruction into a full 64 bit word. As part of this predecoding the S bit of each instruction is expanded to a full 3 bit field 000, 001, ..., 111, which provides the explicit binary group number of the instruction. In other words, the predecoder, by expanding the S bit to a three bit sequence explicitly provides information that the instruction group 000 must execute before instruction group 010, although both groups would have all instructions within the group have S bits set to 0. Because of the frame rules for sequencing groups, these group numbers correspond to the order of issue of the groups of instructions. Group 0 (000) will be issued first, Group 1 (001), if present, will be issued second, Group 2 (010) will be issued third. Ultimately, Group 7 (111), if present, will be issued last. At the time of predecoding of each instructions, the S value of the last word in the frame, which belongs to the last group in the frame to issue, is stored in the tag field for that line in the cache, along with the 19 bit real address and a valid bit. The valid bit is a bit that specifies whether the information in that line in the cache is valid. If the bit is not set to "valid", there cannot be a match or "hit" on this address. The S value from the last instruction, which S value is stored in the tag field of the line in the cache, provides a "countdown" value that can be used to know when to increment to the next cache line.

[0042] As another part of the predecoding process, a new 4 bit field prefix is added to each instruction giving the explicit pipe number of the pipeline to which that instruction will be routed. The use of four bits, rather than three allows the system to be later expanded with additional pipelines. Thus, at the time an instruction is supplied from the predecoder to the instruction cache, each instruction will have the format shown in Figure 6. As shown by Figure 6, bits 0 to 56 provide 57 bits for the instruction, bits 57, 58 and 59 form the full 3 bit S field, and bits 60-63 provide the 4 bit P field.

[0043] Figure 7 illustrates the operation of the predecoder in transferring a frame from memory to the instruction cache. In the upper portion of Figure 7, the frame is shown with a hypothetical four groups of instructions. The first group consists of a single instruction, the sec-

ond group of three instructions, and each of the third and fourth groups of two instructions. As described, instruction is 32 bits in length and include an S bit to separate the groups. The predecoder decodes the instruction shown in the upper portion of Figure 7 into the instruction shown in the lower portion of Figure 7. As shown, the instructions are expanded to 64 bit length, with each instruction including a 4 bit identification of the pipeline to which the instruction is to be assigned, and the expanded group field to designate the groups of instructions that can be executed together. For illustration, hypothetical pipeline tags have been applied. Additionally, the predecoder examines each frame for the minimum number of clocks required to execute the frame, and that number is appended to the address tag 45 for the line. The address tag consists of bits provided for the real address for the line, 1 bit to designate the validity of the frame, and 3 bits to specify the minimum time in number of clock cycles, for that frame to issue. The number of clocks for the frame to issue is determined by the group identification number of the last word in the frame. At this stage, the entire frame shown in the lower portion of Figure 7 is present in the instruction cache.

[0044] It may be desirable to implement the system of this invention on computer systems that already are in existence and therefore have instruction structures that have already been defined without fields for the group information, pipeline information, or both. In this case in another embodiment of this invention the group and pipeline information is supplied on a different clock cycle, then combined with the instructions in the cache. Such an approach can be achieved by adding a "no-op" instruction with fields that identify which instructions are in which group, and identify the pipeline for execution of the instruction, or by supplying the information relating to the parallel instructions in another manner. It therefore should be appreciated that the manner in which the data arrives at the crossbar to be processed is somewhat arbitrary. We use the word "associated" herein to designate the concept that the pipeline and group identifiers are not required to have a fixed relationship to the instruction words. That is, the pipeline and group identifiers need not be embedded within the instructions themselves as shown in Figure 7. Instead they may arrive from another means, or on a different cycle.

[0045] Figure 8 is a simplified diagram illustrating the secondary cache, the predecoder, and the instruction cache. This drawing, as well as Figures 9, 10 and 11, are used to explain the manner in which the instructions tagged with the P and S fields are routed to their designated instruction pipelines.

[0046] In Figure 8 instruction frames are fetched in a single transfer across a 256 bit (32 byte) wide path from a secondary cache 50 into the predecoder 60. As explained above, the predecoder expands each 32 bit instruction in the frame to its full 64 bit wide form and prefixes the P and S fields. After predecoding the 512 bit wide instruction is transferred into the primary instruc-

tion cache 70. At the same time, tag is placed into the tag field 74 for that line.

[0047] The instruction cache operates as a conventional physically-addressed instruction cache. In the example depicted in Figure 8, the instruction cache will contain 512 bit fully-expanded instruction frames of eight instructions each organized in two compartments of 256 lines.

[0048] Address sources for the instruction cache arrive at a multiplexer 80 that selects the next address to be fetched. Because instructions are always machine words, the low order two address bits <1:0> of the 32 bit address field supplied to multiplexer 80 are discarded. These two bits designate byte and half-word boundaries. Of the remaining 30 bits, the next three low order address bits <4:2>, which designate a particular instruction word in a frame, are sent directly via bus 81 to the associative crossbar (explained in conjunction with subsequent figures). The next low eight address bits <12:5> are supplied over bus 82 to the instruction cache 70 where they are used to select one of the 256 lines in the instruction cache. Finally, the remaining 19 bits of the virtual address <31:13> are sent to the translation lookaside buffer (TLB) 90. The TLB translates these bits into the high 19 bits of the physical address. The TLB then supplies them over bus 84 to the instruction cache. In the cache they are compared with the tag of the selected line, to determine if there is a "hit" or a "miss" in the instruction cache.

[0049] If there is a hit in the instruction cache, indicating that the addressed instruction is present in the cache, then the selected frame containing the addressed instruction is transferred across the 512 bit wide bus 73 into the associative crossbar 100. The associative crossbar 100 then dispatches the addressed instruction, with the other instructions in its group, if any, to the appropriate pipelines over buses 110, 111, ..., 117. Preferably the bit lines from the memory cells containing the bits of the instruction are themselves coupled to the associative crossbar. This eliminates the need for numerous sense amplifiers, and allows the crossbar to operate on the lower voltage swing information from the cache line directly, without the normally intervening driver circuit to slow system operation.

[0050] Figure 9 is a block diagram illustrating in more detail the frame selection process. As shown, bits <4:2> of the virtual address are supplied directly to the associative crossbar 100 over bus 81. Bus 81, as explained above will preferably include a pair of conductors, the bit lines, for each data bit in the field. Bits <12:5> supplied over bus 82 are used to select a line in the instruction cache. The remaining 19 bits, translated into the 19 high order bits <31:13> of physical address, are used to compare against the tags of the two selected lines (one from each compartment of the cache) to determine if there is a hit in either compartment. If there is a hit, the two 512 bit wide frames are supplied to multiplexer 120. The choice of which line is ultimately sup-

plied to associative crossbar 100 depends upon the real address bits <31:13> that are compared by comparators 125. The output from comparators 125 thus selects the appropriate frame for transfer to the crossbar 100.

[0051] Figure 10 illustrates in more detail the group select function of the associative crossbar. A 512 bit wide register 130, preferably formed by the SRAM cells in the instruction cache contains the frame of the instructions to be issued. For the purposes of illustration, register 130 is shown as containing a frame having three groups of instructions, with Group 0 including word's W0, W1 and W2; Group 1 containing words W3, W4 and W5; and Group 2 containing words W6 and W7. For illustration, the instructions in Group 0 are to be dispatched to pipelines 1, 2 and 3; the instructions in Group 1 to pipelines 1, 3 and 6; and the instructions in Group 2 to pipelines 1 and 6. The three S bits (group identification field) of each instruction in the frame are brought out to an 8:1 multiplexer 140 over buses 131, 132, 133, ..., 138. The S field of the next group of instructions to be executed is present in a 3 bit register 145. As shown in Figure 10, the hypothetical contents of register 145 are 011. These bits have been loaded into register 145 using bus 81 described in conjunction with Figure 9. Multiplexer 140 then compares the value in this register against the contents of the S field in each of the instruction words. If the two values match, the appropriate decoder 150 is enabled, permitting the instruction word to be processed on that clock cycle. If the values do not match, the decoder is disabled and the instruction words are not processed on that clock cycle. In the example depicted in Figure 10, the contents of register 145 match the S field of the Group 1 instructions. The resulting output, supplied over bus 142, is communicated to S register 144 and then to the decoders via bus 146. The S register contents enable decoders 153, 154 and 155, all of which are in Group 001. As will be shown in Figure 11, this will enable these instructions W3, W4 and W5 to be sent to the pipelines for processing.

[0052] Figure 11 is a block diagram illustrating the group dispatching of the instructions in the group to be executed. The same registers are shown across the upper portion of Figure 11 as in the lower portion of Figure 10. As shown in Figure 11, the crossbar switch itself consists of two sets of crossing pathways. In the horizontal direction are the pipeline pathways 180, 181, ..., 187. The vertical direction are the instruction word paths 190, 191, ..., 197. Each of these pipeline and instruction pathways is itself a bus for transferring the instruction word. Each horizontal pipeline pathway is coupled to a pipeline execution unit 200, 201, 202, 207. Each of the vertical instruction word pathways 190, 191, ..., 197 is coupled to an appropriate portion of register 130 (Figure 10).

[0053] The decoders 170, 171, ..., 177 associated with each instruction word pathway receive the 4 bit pipeline code from the instruction. Each decoder, for example decoder 170, provides as output eight 1 bit con-

control lines. One of these control lines is associated with each pipeline pathway crossing of that instruction word pathway. Selection of a decoder as described with reference to Figure 10 activates the output bit control line corresponding to that input pipe number. This signals the crossbar to close the switch between the word path associated with that decoder and the pipe path selected by that bit line. Establishing the cross connection between these two pathways causes a selected instruction word to flow into the selected pipeline. For example, decoder 173 has received the pipeline bit for word W3. Word W3 has associated with it pipeline path 1. The pipeline path 1 bits are decoded to activate switch 213 to supply instruction word W3 to pipeline execution unit 201 over pipeline path 181. In a similar manner, the identification of pipeline path 3 for decoder D4 activates switch 234 to supply instruction word W4 to pipeline path 3. Finally, the identification of pipeline 6 for word W5 in decoder D5 activates switch 265 to transfer instruction word W5 to pipeline execution unit 206 over pipeline pathway 186. Thus, instructions W3, W4, and W5 are executed by pipes 201, 203 and 206, respectively.

[0054] The pipeline processing units 200, 201, ..., 207 shown in Figure 11 can carry out desired operations. In a preferred embodiment of the invention, each of the eight pipelines first includes a sense amplifier to detect the state of the signals on the bit lines. In one embodiment the pipelines include first and second arithmetic logic units; first and second floating point units; first and second load units; a store unit and a control unit. The particular pipeline to which a given instruction word is dispatched will depend upon hardware constraints as well as data dependencies.

[0055] Figure 12 is an example of a frame and how it will be executed by the pipeline processors 200-207 of Figure 11. As shown in Figure 12, the frame includes three groups of instructions. The first group, with group identification number 0, includes two instructions that can be executed by the arithmetic logic unit, a load instruction and a store instruction. Because all these instructions have been assigned the same group identification number by the compiler, all four instructions can execute in parallel. The second group of instructions consists of a single load instruction and two floating point instructions. Again, because each of these instructions has been assigned "Group 1", all three instructions can be executed in parallel. Finally, the last instruction word in the frame is a branch instruction that, based upon the compiler's decision, must be executed last.

[0056] Figure 13 illustrates the execution of the instructions in the frame shown in Figure 12. As shown, during the first clock the Group 0 instructions execute, during the second clock the load and floating point instructions execute, and during the third clock the branch instruction executes. To prevent groups from being split across two instruction frames, an instruction frame may be only partially filled, where the last group is too large to fit entirely within the remaining space of the frame.

[0057] Figure 14 is a diagram illustrating another embodiment of the associative crossbar. In Figure 14 nine pipelines 0 - 8 are shown coupled to the crossbar. The three bit program counter PC points to one of the instructions in the frame, in combination with the set of 8 group identification bits for the frame, indicating the group affiliation of each instruction, are used to enable a subset of the instructions in the frame. The enabled instructions are those at or above the address indicated by the PC that belong to the current group.

[0058] The execution ports that connect to the pipelines specified by the pipeline identification bits of the enabled instructions are then selected to multiplex out the appropriate instructions from the current frame. If one or more of the pipelines is not ready to receive a new instruction, as set of hold latches at the output of the execution ports prevents any of the enabled instructions from issuing until the "busy" pipeline is free. Otherwise the instructions pass transparently through the hold latches into their respective pipelines. Accompanying the output of each port is a "port valid" signal that indicates whether the port has valid information to issue to the hold latch.

[0059] Figure 15 is a diagram illustrating the group select function in further detail. This figure illustrates the mechanism used to enable an addressed group of instructions within a frame. The program counter is first decoded into a set of 14 bit signals. Seven of these signals are combined with the eight group identifiers of the current frame to determine whether each of the seven instructions, 11 to 17, is or is not the start of a later group. This information can then be combined with the other 7 bit signals from the PC decoder to determine which of the eight instructions in the frame should be enabled. Using the pipeline identifying field each enabled instruction can be combined with the other 7 bit signal to determine which of the eight instructions in the frame should be enabled. Each such enabled instruction can then signal the execution port, as determined by the pipeline identifier, to multiplex out the enabled instruction. Thus, if I2 is enabled, and the pipeline code is 5, the select line from I2 to port 5 is activated, causing I2 to flow to the hold latch at pipe 5.

[0060] Because the instructions that start later groups are known, the system can decide easily which instructions start the next group. This information is used to update the PC to the address of the next group of instructions. If no instruction in the frame begins the next group, i.e., the last instruction group has been dispatched to the pipelines, a flag is set. The flag causes the next frame of instructions to be brought into the crossbar. The PC is then reset to 10. Shown in the figure is an exemplary sequence of the values that the PC, the instruction enable bits and the next frame flag take on over a sequence of eight clocks extending over two frames.

[0061] The processor architecture described above provides many unique advantages to a system using

this invention. The system described is extremely flexible, enabling instructions to be executed sequentially or in parallel, depending entirely upon the "intelligence" of the compiler. As compiler technology improves, the described hardware can execute programs more rapidly, not being limited to any particular frame width, number of instructions capable of parallel execution, or other external constraints. Importantly, the associative crossbar aspect of this invention relies upon the content of the message being decoded, not upon an external control circuit acting independently of the instructions being executed. In essence, the associative crossbar is self directed. In the preferred embodiment the system is capable of a parallel issue of up to eight operations per cycle. For amore complete description of the associative crossbar, see EP-A-0 652 509.

[0062] Although the foregoing has been a description of the preferred embodiment of the invention, it will be apparent to those of skill in the art that numerous modifications and variations may be made to the invention without departing from the scope as described therein. For example, arbitrary numbers of pipelines, arbitrary numbers of decoders, and different architectures may be employed, yet rely upon the system we have developed.

Claims

1. A method for operating a computing system comprising:

fetching a first frame of instructions from an instruction memory, the first frame of instructions comprising up to eight instructions, each instruction comprising instruction data and a grouping bit, wherein grouping bits of the instructions are indicative of groups of instructions from the first frame of instructions, wherein the groups of instructions comprises at least one group of instructions and comprises at most eight groups of individual instructions, wherein a group of instructions are issued separately from other groups of instructions within the first frame of instructions, wherein groups of instructions are issued from left-to-right from the first frame of instructions, and wherein instructions within a group of instructions are issued in parallel;

issuing instructions in a first group of instructions from the first frame of instructions to functional units appropriate for the instructions in the first group of instructions in response to grouping bits of the instructions in the first group of instructions and in response to a mapping of the instructions in the first group of instructions to functional units, wherein the mapping is determined in response to at least a por-

tion of instruction data in each instruction in the first group of instructions from the first frame of instructions;

wherein the groups of instructions from the first frame of instructions are specified at compile time.

2. The method of claim 1 comprising:

fetching a second frame of instructions from the instruction memory, the second frame of instructions comprising up to eight instructions, each instruction comprising instruction data and a grouping bit, wherein grouping bits of the instructions are indicative of groups of instructions from the second frame of instructions, wherein the groups of instructions comprises at least one group of instructions and comprises at most eight groups of individual instructions, wherein a group of instructions are issued separately from other groups of instructions within the second frame of instructions, wherein groups of instructions are issued from left-to-right from the second frame of instructions, and wherein instructions within a group of instructions are issued in parallel; and
issuing instructions in a first group of instructions from the second frame of instructions to functional units appropriate for the instructions in the first group of instructions in response to grouping bits of the instructions in the first group of instructions and in response to a mapping of the instructions in the first group of instructions to functional units, wherein the mapping is determined in response to at least a portion of instruction data in each instruction in the first group of instructions from the second frame of instructions;

wherein instructions from the first frame of instructions are issued before fetching the second frame of instructions from the instruction memory; and

wherein the groups of instructions from the second frame of instructions are specified at compile time.

3. The method of claim of claim 2 wherein a 256 bits wide first frame of instructions and 32 bits wide instructions are used.
4. The method of any of the above claims wherein a first frame of instructions with fewer than then eight instructions is used.
5. The method of any of the above claims further comprising:

- issuing instructions in a second group of instructions from the first frame of instructions to functional units appropriate for the instructions in the second group of instructions in response to grouping bits of the instructions in the second group of instructions and in response to a mapping of the instructions in the second group of instructions to functional units, wherein the mapping is determined in response to at least a portion of instruction data in each instruction in the second group of instructions from the first frame of instructions.
6. The method of claim 5 wherein a functional unit appropriate for an instruction in the second group of instructions from the first frame of instruction comprises a floating point unit; and wherein the second group of instructions from the first frame of instructions includes a floating point instruction.
 7. The method of claims any of the above claims wherein a first frame of instructions with exactly eight 32 bit-wide instructions is used.
 8. The method of any of the above claims wherein three or fewer groups of instructions are packed within the first frame of instructions thereby reducing instruction code size.
 9. The method of any of the above claims wherein issuing instructions in a first group of instructions from the first frame of instructions further comprises decoding the instructions in the first group of instructions before issuance to functional units appropriate for the instructions in the first group of instructions from the first frame of instructions.
 10. The method of any of the above claims wherein groups of instructions are not split across the first frame of instructions and the second frame of instructions.
 11. The method of any of the above claims wherein the groups of instructions from the first frame of instructions and the groups of instructions from the second frame of instructions are specified at compile time in response to data dependency checking of the instructions in the first frame of instructions and in response to data dependency checking of the instructions in the second frame of instructions.
 12. The method of any of the above claims wherein the first group of instructions from the first frame of instructions includes two add instructions that are issued in parallel to two arithmetic logic units.
 13. The method of any of the above claims comprising including in the first group of instructions from the first frame of instructions a load instruction that operates upon a register.
 14. The method of any of the above claims wherein a branch instruction is used as an instruction in the first frame of instructions.
 15. A processor comprising:
 - a plurality of functional units;
 - an instruction memory configured to store a first frame of instructions, the first frame of instructions comprising up to eight instructions, each instruction comprising instruction data and a grouping bit, wherein grouping bits of the instructions are indicative of variable length groups of instructions from the first frame of instructions, wherein the variable length groups of instructions comprises at least one group of instructions and comprises at most eight groups of individual instructions, wherein a group of instructions are dispatched separately from other groups of instructions within the first frame of instructions, wherein groups of instructions are dispatched from left-to-right from the first frame of instructions, and
 - wherein instructions within a group of instructions are dispatched in parallel; and
 - an instruction dispatching unit coupled to the instruction memory and the plurality of functional units, the instruction dispatching unit configured to dispatch instructions in a first group of instructions from the first frame of instructions to functional units appropriate for the instructions in the first group of instructions in response to grouping bits of the instructions in the first group of instructions and in response to a mapping of the instructions in the first group of instructions to functional units, wherein the mapping is determined in response to at least a portion of instruction data in each instruction in the first group of instructions from the first frame of instructions; and
 - wherein the groups of instructions from the first frame of instructions are determined at compile time.
 16. The processor of claim 15 wherein the first frame of instructions is 256 bits-wide; and wherein the instructions in the first frame of instructions comprise 32 bit words.
 17. The processor of any of the above claims wherein the first frame of instructions comprises less than eight 32 bit-wide instructions and a no operation.
 18. The processor of any of the above claims wherein the plurality of functional units comprises a function-

al unit configured to perform floating-point operations.

19. The processor of any of the above claims wherein the plurality of functional units comprises a functional unit configured to perform a load operation and a store operation. 5
20. The processor of any of the above claims wherein the plurality of functional units comprises a functional unit configured to perform branch instructions. 10
21. The processor of any of the above claims wherein at least three groups of instructions are packed within the first frame of instructions thereby saving memory space. 15
22. The processor of any of the above claims wherein the instruction dispatching unit is also configured to fetch the first frame of instructions and configured to decode the instructions in the first group of instructions before dispatch to functional units appropriate for the instructions in the first group of instructions from the first frame of instructions. 20
25
23. The processor of any of the above claims wherein groups of instructions to be dispatched in parallel are not split across frames of instructions.
24. The processor of any of the above claims wherein at least four groups of instructions are packed within the first frame of instructions thereby reducing memory bandwidth. 30
25. The processor of any of the above claims wherein the groups of instructions from the first frame of instructions are specified at compile time by the grouping bits of the instructions in the first frame of instructions in response to data dependency checking of the instructions in the first frame of instructions. 35
40
26. The processor of any of the above claims wherein the instructions in the first instruction frame are stored in little-endian format. 45
27. The processor of any of the above claims wherein the plurality of functional units include functional units configured to perform arithmetic and logic operations. 50
28. An apparatus comprises:
 - a memory configured to store a plurality of instruction data; and 55
 - a processor coupled to the memory comprising:
 - a memory controller configured to receive

a first set of instruction data from the memory;
 a plurality of functional units;
 an instruction memory configured to store a first frame of instructions in response to the first set of instruction data from the memory, the first frame of instructions comprising up to eight instructions, each instruction comprising instruction data and a grouping bit, wherein grouping bits of the instructions are indicative of groups of instructions of variable lengths from the first frame of instructions, wherein the groups of instructions comprises at least one group of instructions and comprises at most eight groups of instructions, wherein a number of instructions in the groups of instructions comprises at least one instruction and up to eight instructions; wherein a group of instructions are dispatched separately from other groups of instructions within the first frame of instructions, wherein groups of instructions are dispatched from left-to-right from the first frame of instructions, and wherein instructions within a group of instructions are dispatched in parallel; and
 an instruction dispatching unit coupled to the instruction memory and coupled to the plurality of functional units, the instruction dispatching unit configured to fetch the first frame of instructions from the instruction memory and configured to dispatch instructions in a first group of instructions from the first frame of instructions to functional units appropriate for the instructions in the first group of instructions in response to grouping bits of the instructions in the first group of instructions and in response to a mapping of the instructions in the first group of instructions to appropriate functional units, wherein the mapping is determined in response to at least a portion of instruction data in each instruction in the first group of instructions from the first frame of instructions;

wherein the groups of instructions from the first frame of instructions are determined at compile time.

29. The apparatus of claim 28 wherein the first frame of instruction comprises 256 bits and the instructions in the first frame of instructions comprises 32 bits.
30. The apparatus of any of the above claims wherein the memory controller is configured

to receive a second set of instruction data from the memory;

wherein the instruction memory is also configured to store a second frame of instructions in response to the second set of instruction data from the memory, the second frame of instructions comprising up to eight instructions, each instruction comprising instruction data and a grouping bit, wherein grouping bits of the instructions are indicative of groups of instructions of variable lengths from the second frame of instructions, wherein the groups of instructions comprises at least one group of instructions and comprises at most eight groups of instructions, wherein a number of instructions in the groups of instructions comprises at least one instruction and up to eight instructions; wherein a group of instructions are dispatched separately from other groups of instructions within the second frame of instructions, wherein groups of instructions are dispatched from left-to-right from the first frame of instructions, and wherein instructions within a group of instructions are dispatched in parallel; and wherein the instruction dispatching unit is also configured to fetch the second frame of instructions from the instruction memory and configured to dispatch instructions in a first group of instructions from the second frame of instructions to functional units appropriate for the instructions in the first group of instructions in response to grouping bits of the instructions in the first group of instructions and in response to a mapping of the instructions in the first group of instructions to appropriate functional units, wherein the mapping is determined in response to at least a portion of instruction data in each instruction in the first group of instructions from the second frame of instructions;

wherein the first group of instructions from the first frame of instructions and the first group of instructions in the second wide frame of instructions are determined at compile time.

31. The apparatus of any of the above claims wherein the first frame of instructions comprises a no operation instruction.

32. The apparatus of any of the above claims wherein the plurality of functional units comprises eight functional units including a floating-point unit.

33. The apparatus of any of the above claims wherein the plurality of functional units also includes a functional unit configured to perform a load operation.

34. The apparatus of any of the above claims wherein more than two groups of instructions are packed within the first frame of instructions thereby reducing instruction code size.

35. The apparatus of any of the above claims wherein the instruction dispatching unit is also configured to decode the instructions in the first group of instructions before dispatch to the appropriate functional units.

36. The apparatus of any of the above claims wherein the groups of instructions from the first frame of instructions are specified at compile time by the grouping bits of the instructions in the first frame of instructions in response to data dependency checking of the instructions in the first frame of instructions.

37. The apparatus of any of the above claims wherein the groups of instructions from the first frame of instructions comprises the first group of instructions, a second group of instructions, and a third group of instructions; and wherein a number of instructions in the first group of instructions is different from a number of instructions in the second group of instructions.

38. The apparatus of any of the above claims wherein the groups of instructions from the first frame of instructions comprise four groups of instructions; and wherein a number of instructions in the first group of instructions is different from a number of instructions in a fourth group of instructions.

39. The apparatus of any of the above claims wherein a number of instructions in the first group of instructions is equal to a number of instructions in the second group of instructions.

40. The apparatus of any of the above claims wherein the number of instructions in the first group of instructions is exactly one instruction.

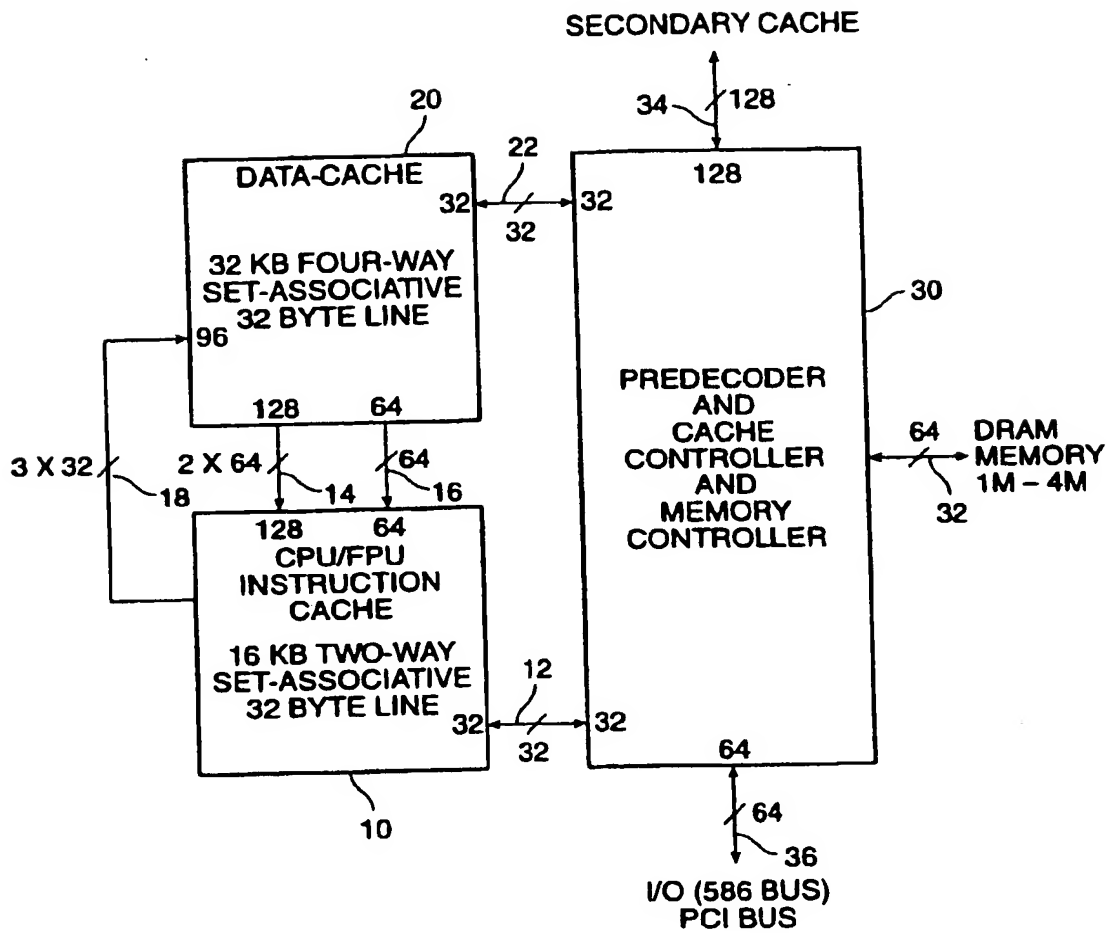


FIG. 1



FIG. 2



FIG. 3

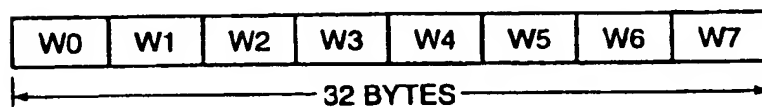


FIG. 4

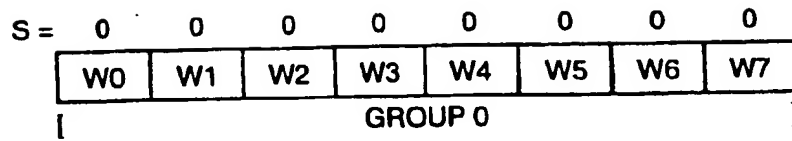


FIG. 5A

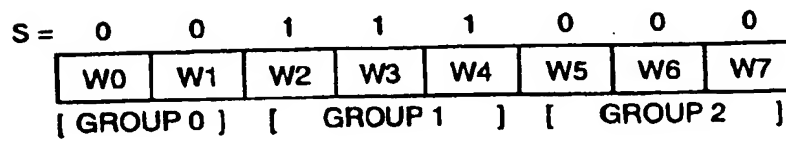


FIG. 5B

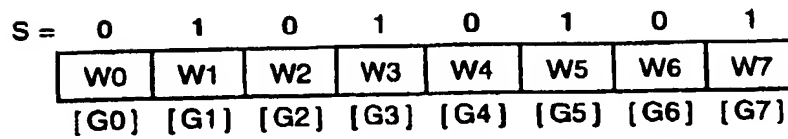


FIG. 5C



FIG. 6

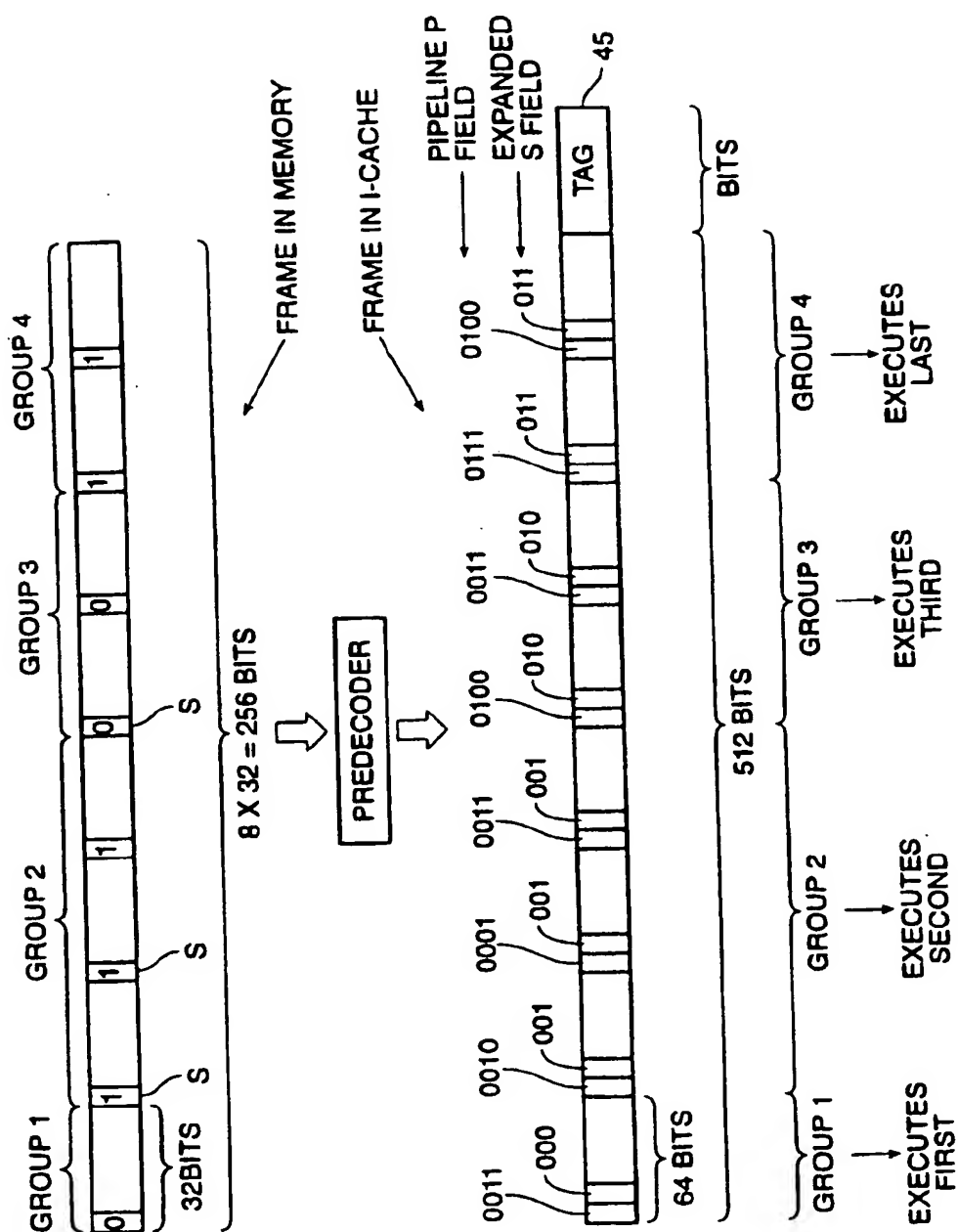


FIG. 7

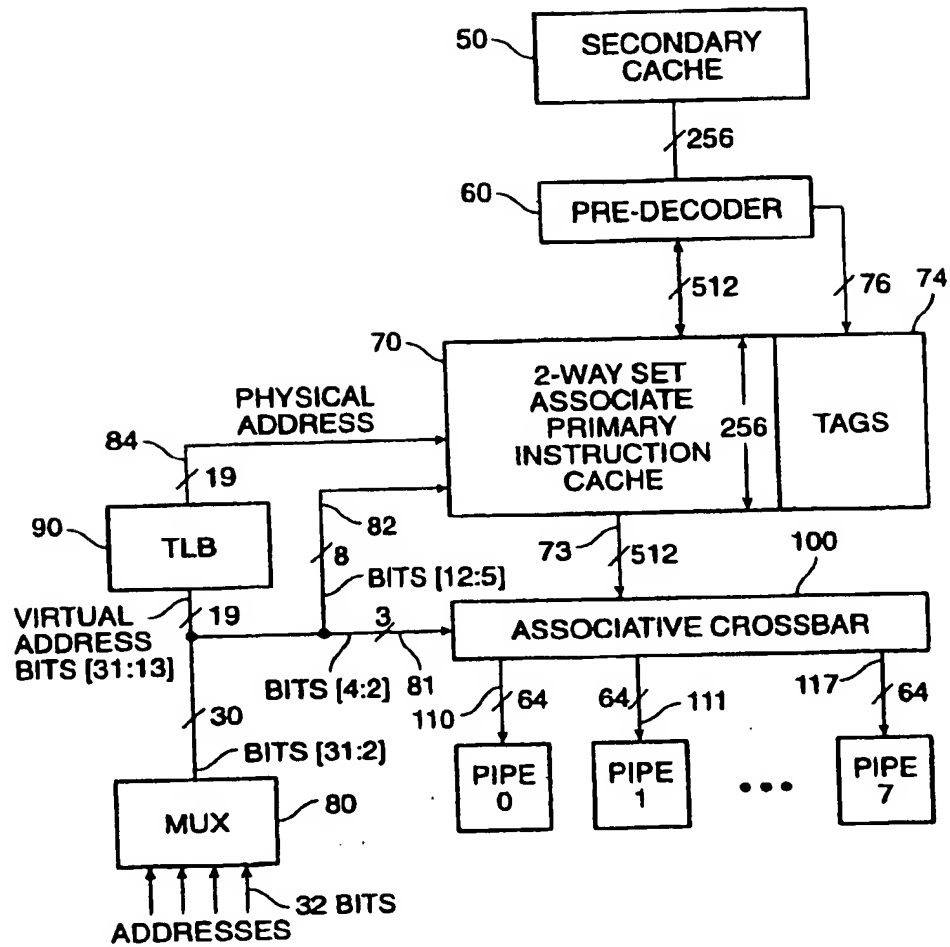


FIG. 8

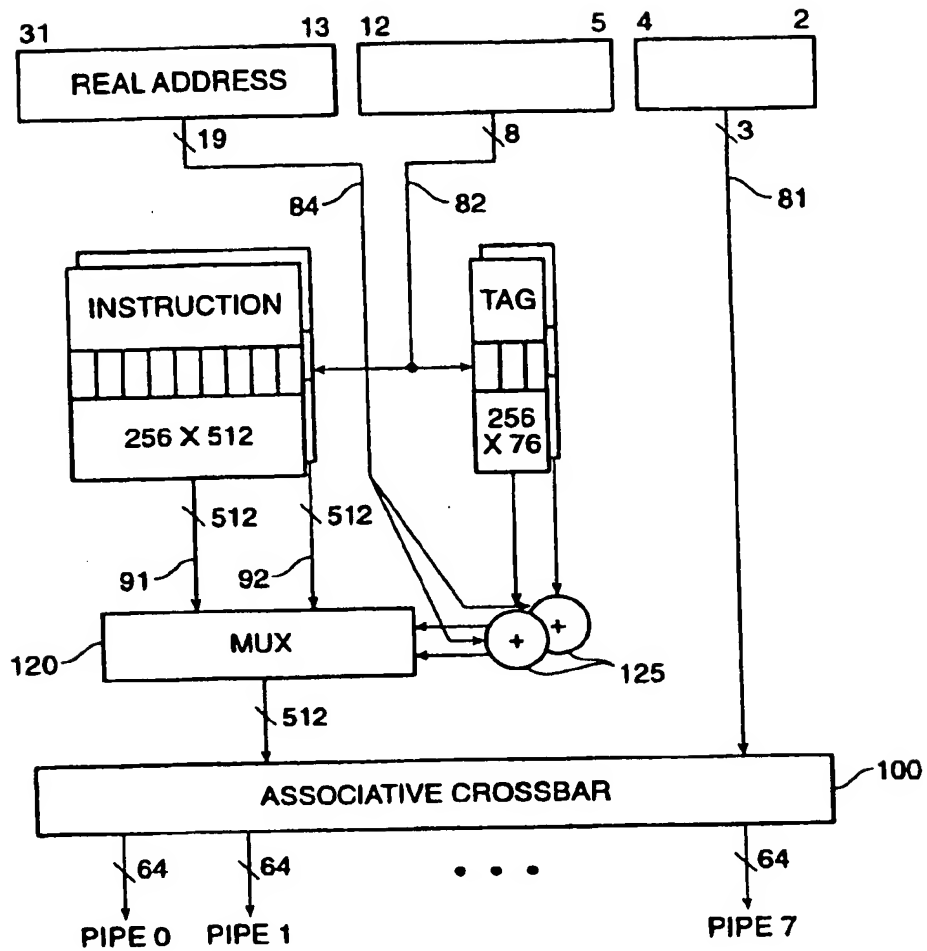


FIG. 9

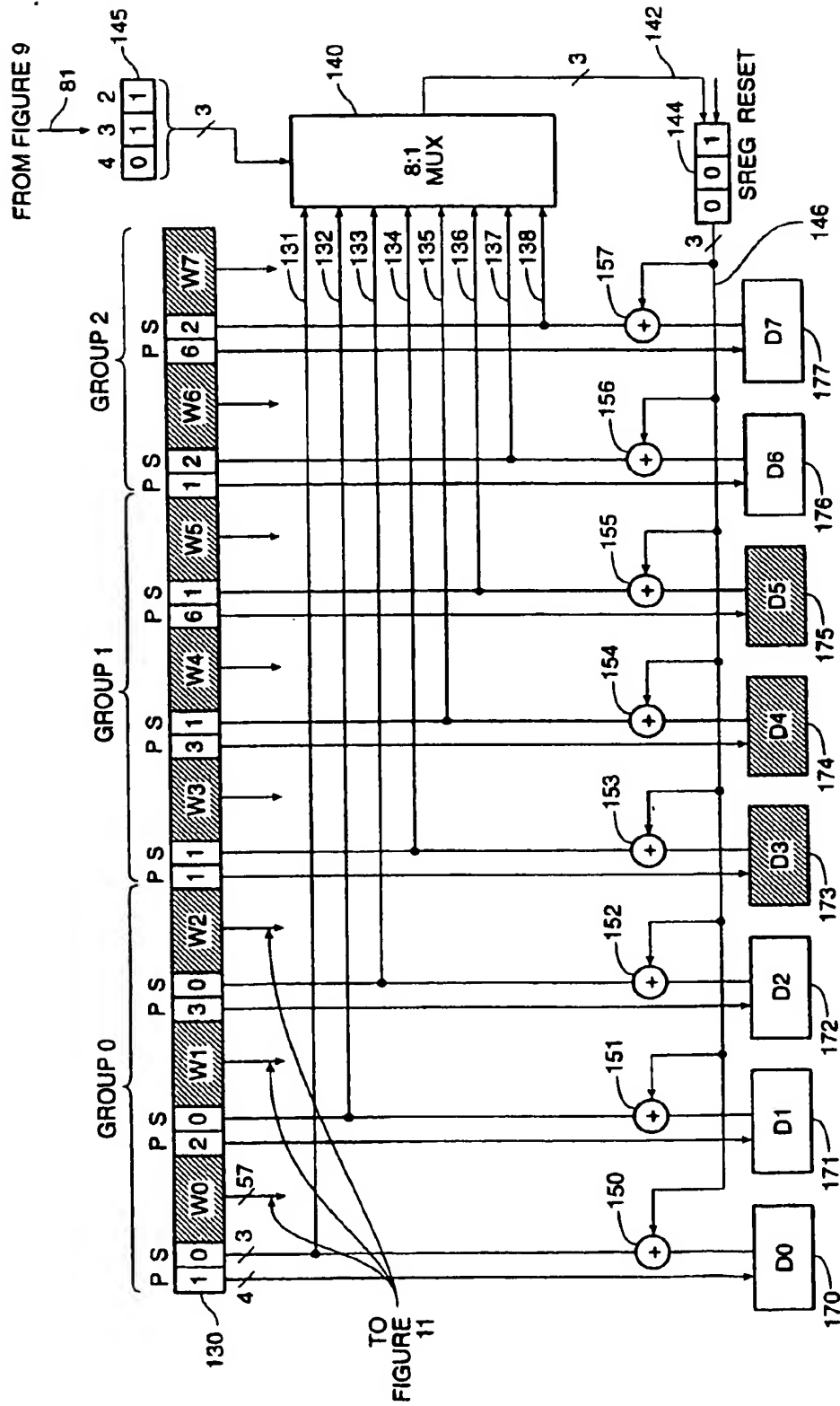


FIG. 10

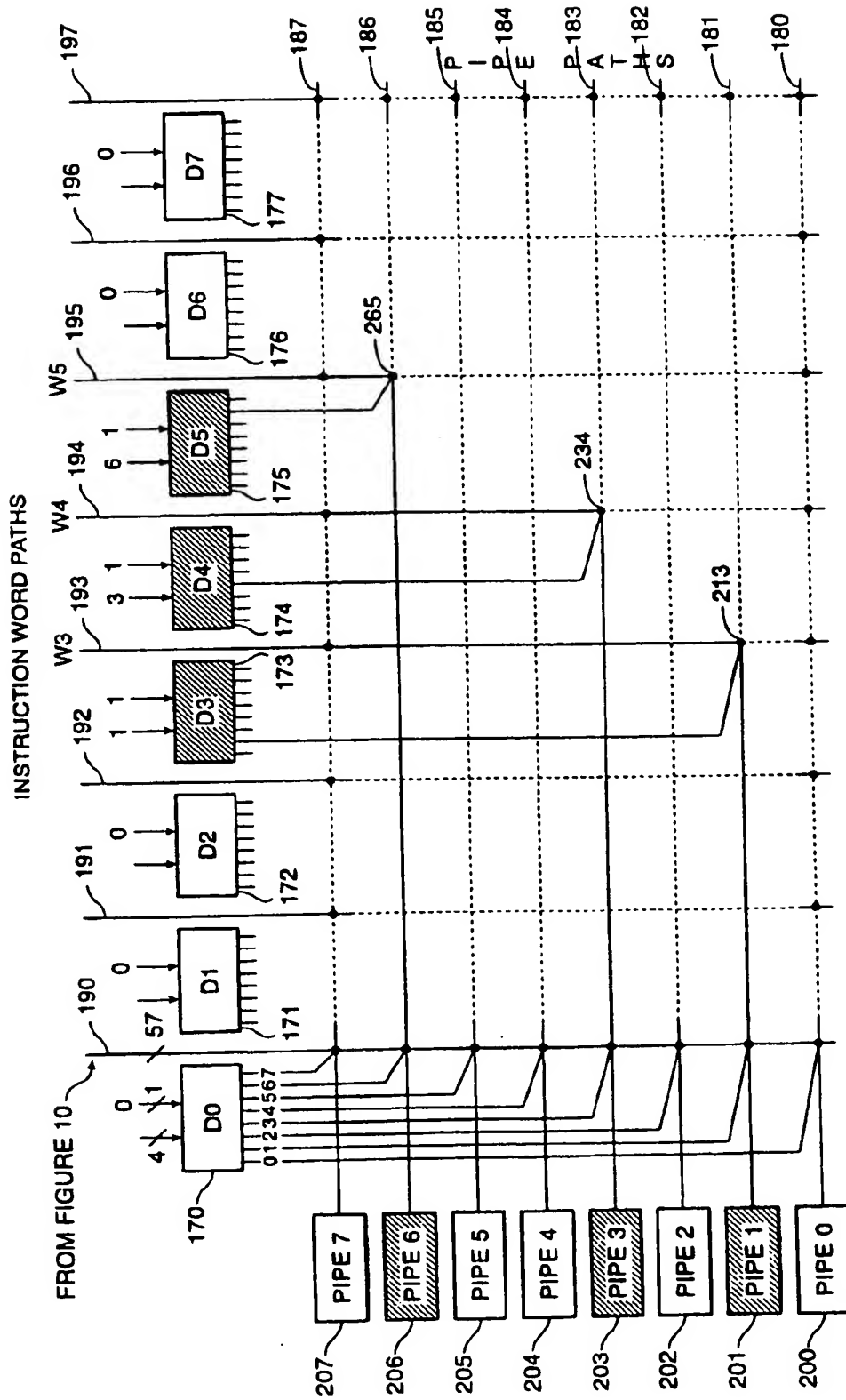


FIG. 11

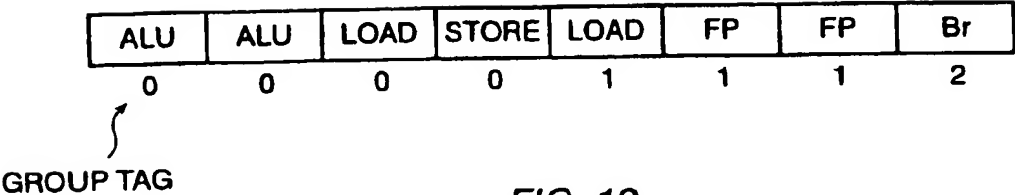


FIG. 12

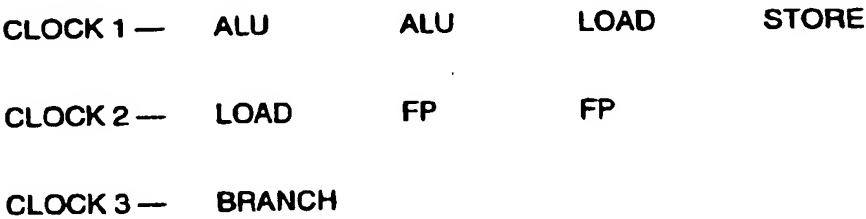


FIG. 13

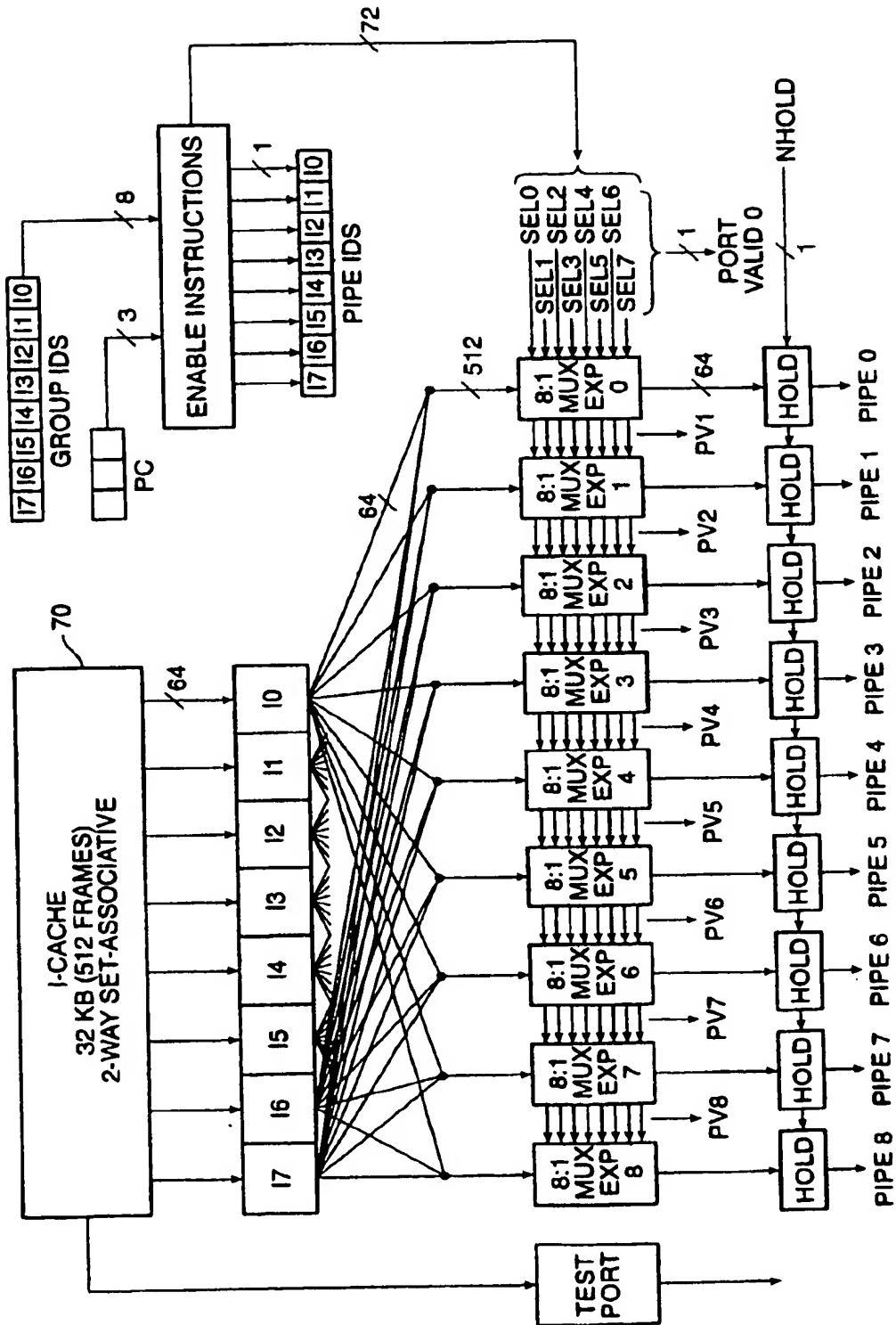


FIG. 14

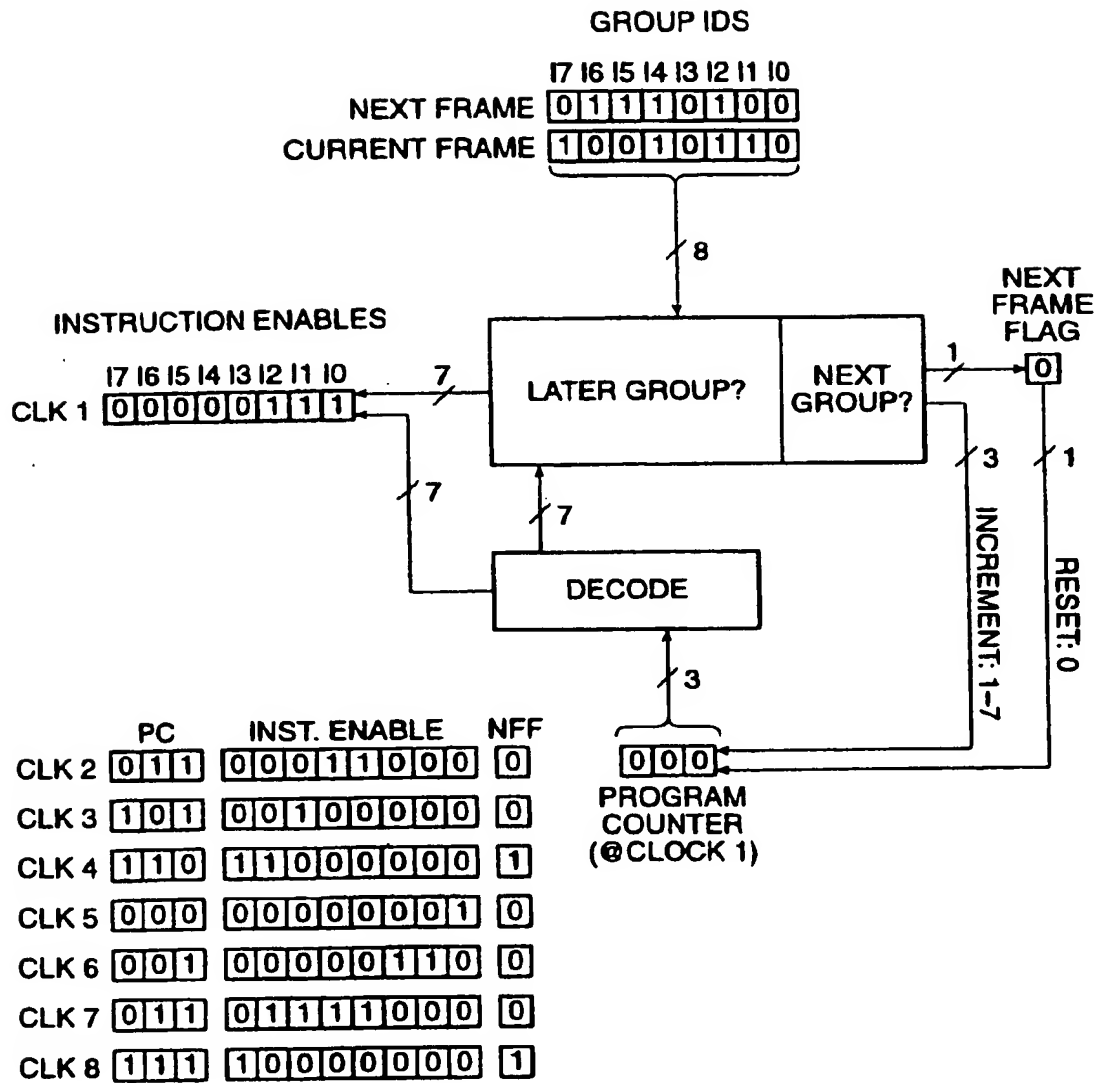


FIG. 15